

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 60-160475

(43)Date of publication of application : 22.08.1985

(51)Int.Cl.

G06F 15/40

(21)Application number : 59-015184

(71)Applicant : HITACHI LTD

(22)Date of filing : 01.02.1984

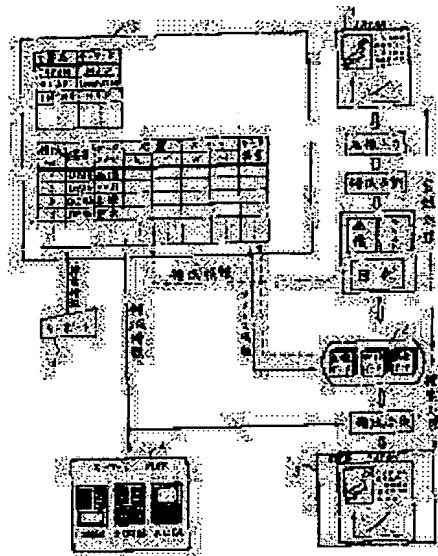
(72)Inventor : HINO MASATOSHI  
TABATA KUNIAKI

## (54) METHOD FOR RETRIEVING DOCUMENT FILE

### (57)Abstract

**PURPOSE:** To improve operability and retrieval efficiency by dividing documents in areas where the same quality data exist at the time of registration and using it as structural information of the documents for retrieval.

**CONSTITUTION:** A method divides documents into areas containing only the same quality data such as a text a picture, etc. and a graph of a sentence 1 of right/ left, to store them respectively in a document file 2 while registering in a file. At the same time, the method registers a document name, retrieval key word together with an area No. for each area, data type, position, size, data No. as structural information of the documents from a keyboard 6 to register them in a retrieval data base 5. When retrieving a desired document from the file, the device, if there exist plural documents coincident with the designated retrieval information, makes them as each of document structures to display them on a display by the patterns in correspondence to their data of each area.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japan Patent Office

BEST AVAILABLE COPY

⑬ 日本国特許庁(JP)

⑭ 特許出願公開

⑯ 公開特許公報(A)

昭60-160475

⑮ Int. Cl.<sup>4</sup>

識別記号

庁内整理番号

⑰ 公開 昭和60年(1985)8月22日

G 06 F 15/40

Z-6913-5B

審査請求 未請求 発明の数 1 (全7頁)

⑱ 発明の名称 文書ファイル検索方式

⑲ 特 願 昭59-15184

⑳ 出 願 昭59(1984)2月1日

㉑ 発 明 者 樋 野 匡 利 川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

㉒ 発 明 者 田 畑 邦 晃 川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

㉓ 出 願 人 株式会社日立製作所 東京都千代田区神田駿河台4丁目6番地

㉔ 代 理 人 弁理士 高橋 明夫 外1名

## 明 細 書

発明の名称 文書ファイル検索方式

## 特許請求の範囲

1. 文書データを蓄積する文書ファイルを備えたファイル検索システムにおいて、上記文書ファイルに文書データを蓄積するときは、テキスト、画像、図表等の同質データを含む領域に分割して蓄積し、指定情報を入力して文書を検索するときは、上記指定情報に合致する文書が複数件存在するならば、文書を構成する領域を内容に応じた構造パターンで表示し、表示された複数の文書の中から目的の文書を認識することを特徴とする文書ファイル検索方式。
2. 前記内容に応じた構造パターンは、構造表示画面に表示され、該画面中の任意の文書または領域は、文書表示画面により表示されることを特徴とする特許請求の範囲第1項記載の文書ファイル検索方式。
3. 前記内容に応じた構造パターンは、領域に関するデータ・タイプ、位置、大きさ等の情報を

指定することにより表示されることを特徴とする特許請求の範囲第1項または第2項記載の文書ファイル検索方式。

## 発明の詳細な説明

## 〔発明の利用分野〕

本発明は文書ファイル検索方式に関し、特に登録時に文書を同質のデータの存在する領域に分割し、それを文書の構造情報として、検索時に利用するファイル検索方式に関するものである。

## 〔発明の背景〕

従来より用いられている文書ファイル検索方式としては、基本的なものに、(i) 検索すべき文書名を直接指定する方式や、(ii) 分類コード(大分類、中分類等)の指定に応じて表示される候補データの中から選択する方式がある。また、より高度な検索方式としては、(iii) 画像の例示で画像ファイルを検索するQPE(Query-by-Pictorial-Example: 日経コンピュータ1982.5.31 pp40~49参照)、(iv) データ内容から連想した色彩と対応づける方式(特開昭57

-121767号公報参照)、(V)漢字パターン列の特徴パラメータによる検索(電子通信学会研究會資料1E80-120pp91-98大田、その他「漢字パターン列の特徴パラメータによる検索」参照)、および(VI)部分画像や縮小画像を案内画像として用いる方式等が知られている。

以上は、機械によりファイルを検索する方法であるが、一方、通常の人間が機械の力を借りずに行う検索方法を考えると、先ずその文書の内容等に応じて、それがどのファイル、どの書路に存在するかを考慮し、該当するファイルや書路を選び出す(文書ファイルではキーワードによるおおまかな検索がこれに相当する)。次に、その文書のどこに図があるか、あるいは絵や写真があるか等の構造に関する情報を想像しながら、目標とする文書を探し出す。このような文書の構造に関する情報は、元の文書そのものよりデータ量がきわめて少なく、かつ文書の概略構造をディスプレイに高速に表示することが可能であつて、検索情報自体としても使用することができる。

するときは、上記指定情報に合致する文書が複数存在するならば、文書を作成する領域を内容に応じた構造パターンで表示し、表示された複数の文書の中から目的の文書を認識することに特徴がある。

#### 〔発明の実施例〕

以下、本発明の実施例を、図面により説明する。

第2図は、本発明の原理を示す概念図であり、第2図は文書と領域の位置関係を示す図である。

第1図において、1は文書、2は文書ファイル、3は文書表示画面、4は構造表示画面、5は検索データ・ベース、6はキーボードである。

右方に示すように、第1図の上半分は登録処理、下半分は検索処理を表わしている。

先ず、(1)文書をファイルに登録するときには、同質のデータ、例えばテキスト、画像等だけを含む領域に分割して、文書ファイルに蓄積する。このとき、文書名、検索キーワードとともに、各領域のデータ・タイプ、位置、大きさ等を、文書の構造情報として検索データ・ベース5に登録す

光ディスクによる文書画像ファイルの開発や、テキストだけでなく画像や図表を扱うことができる文書処理システムの開発が進んでおり、特に文書から画像領域、テキスト領域、図表領域等を抽出し、各領域に適した情報圧縮を行う方式(情報処理学会第21回全国大会、7H-1、村尾、坂井「文書画像における構造情報の抽出」参照)も提案されているが、まで文書の構造に関する情報を、検索に利用した例は見当たらない。

#### 〔発明の目的〕

本発明の目的は、文書の構造に関する情報を利用し、操作性に優れ検索効率がよい文書ファイルの検索方式を提供することである。

#### 〔発明の概要〕

上記目的を達成するため、本発明の文書ファイル検索方式は、文書データを蓄積する文書ファイルを備えたファイル検索システムにおいて、上記文書ファイルに文書データを蓄積するときは、テキスト、画像、図表等の同質データを含む領域に分割して蓄積し、指定情報を入力して文書を検索

する。また、各領域に対応するデータの文書ファイル中のデータ番号も登録する。

例えば、第1図においては、右方の文書1に示すように、同質のデータ、すなわち×印で示すテキスト、地図等の画像、図表等のみを含む領域に分割して、文書ファイル2にそれぞれ蓄積する。同時に、キーボード6から文書名、検索キーワードとともに、上記領域ごとに領域別、データ・タイプ、位置、大きさ、データ番号を文書の構造情報として、検索データ・ベース5に登録する。登録するとき、文書名とキーワードのみのテーブルと、その他の構造情報を記憶したテーブルの2種類に分けておく。

文書1と領域8の位置関係は、第2図に示すように、文書1の左端からの距離Xと上端からの距離Yで位置を、また高さHと幅Wの寸法で大きさを表わす。また、テーブル中に登録されるデータ番号は、各領域8に対応するデータの文書ファイル2中の番号である。領域分割は、計算機により自動的に行うことも、また人手により行うことも

可能である。検索データ・ベース5は、リレーショナル・データ・ベース等のよく知られた技術で構成される。

次に(II)ファイルから目的の文書を検索するときには、指定された検索情報に合致する文書が複数存在すれば、各文書の構造として、各領域をそのデータに応じたパターンでディスプレイに表示する(これを構造表示画面と呼ぶ)。ただし、検索情報から一意に決まる場合は、該当の文書データを文書ファイルから読み出し、ディスプレイに表示する(これを文書表示画面と呼ぶ)。

また、構造表示画面上の指定した文書の指定領域のデータを、文書ファイルから読み出し文書表示画面に表示することができる。また、構造表示画面上で、目的の文書を探し、当該文書を指定する。これにより、指定の文書データを文書ファイルから読み出して、文書表示画面に表示する。なお、構造表示画面、文書表示画面は、別個のディスプレイを用いてもよいが、同一のディスプレイを用いて時分的に、あるいは画面を2分割し

て表示することもできる。なお、構造表示画面から目的の文書をパターン認識して探索することもできる。

検索方法を、第1図により具体的に説明する。

#### (a) キーワードによる検索

(イ) 目的文書のキーワードをキーボード6から指定する。(ロ) 検索データ・ベース5より指定のキーワードを持つ文書名を求める。(ハ) 指定のキーワードを持つ文書が複数存在する場合、それらの文書の構造情報、すなわちその文書に属する領域のデータ・タイプ、位置、大きさを求め、そのデータ・タイプ、位置、大きさを求め、そのデータ・タイプに応じたパターンで構造表示画面4に表示する。第1図では、「MAP」をキーワードとした、3つの地図が構造表示画面4に表示されている。ただし、文書が一意に決定される場合は、該当の文書を文書ファイル2から読み出して、領域合成し、文書表示画面3に表示する。

(ニ) 構造表示画面4上の文書の領域を指定し、その内容を文書表示画面3に表示することもでき

るので、この表示結果を目的文書の探索に利用できる。(ホ) 構造表示画面4上で、目的の文書をパターン認識により指定する。データ・ベース5より指定された文書の各領域の情報とデータ番号を求めて、文書ファイル2から読み出し、領域合成して文書表示画面3に表示する。

#### (b) 構造情報による検索

(イ) 目的の文書情報、例えば、図が2個存在すること、あるいは右上に画像が存在すること等を、キーボード6から入力する。(ロ) 検索データ・ベース5により、パターン認識の結果、入力された検索情報に合致する文書名を求める。(ハ) 以下は(a)キーワードによる検索の(ハ)以降の処理と同じである。

#### (c) 文書名の直接指定による検索

文書名を直接指定して、目的の文書を文書表示画面3に表示することもできる。

なお、領域分割された結果の各領域のデータは、領域ごとに文書ファイル2に蓄積されるが、このとき、データ・タイプによつて圧縮や蓄積方法を

変えて蓄積することにより、蓄積効率を向上させることもできる。実施例では、文書上の位置、大きさ、文書ファイルのデータ番号が与えられたとき、領域合成により文書表示画面3に表示することができれば、どのような構成であつてもよい。

第3図は、本発明の一実施例を示す文書ファイル検索システムの機能ブロック図である。

第3図において、10はマイクロプロセッサ、20は主メモリ、30はキーボード、40はファイル装置、50は画像プロセッサ、60は画像メモリ、70は画像入力装置、80は画像ディスプレイ、90はデータ・ベース・システムである。

第1図に示す文書ファイル2は、ファイル装置40に存在し、検索データ・ベース5はデータベース・システム90上に存在する。第3図のシステムは、ファイル装置40に文書データを蓄積し、必要に応じて目的の画像データをファイル装置40から読み出し、画像ディスプレイ80に表示する機能を有している。そして、このシステムは、各種のコマンドの指示によつて動作し、コマンド

は主メモリ20上のプログラムとデータにもとづいて、マイクロプロセッサ10により実行される。画像入力装置70は、例えばCCD等により画像を読み取ってシステムに入力するためのものであり、画像プロセッサ50と画像メモリ60は、検索処理等のため領域合成や、画像ディスプレイ80に構造表示画面と文書表示画面を出力するためのプログラムを実行する機能を有するものである。

第4図は、本発明に用いられるコマンドとその機能の説明図であり、第5図は、第4図のコマンドの処理フローチャートである。

STOREのコマンドは、文書の読み込み、領域分割および登録を行う指示を与えるものであり、FINDのコマンドは検索情報に合致する文書を検索し、文書ファイルから読み出して表示する指示を与えるものであり、DELETEのコマンドは文書データを削除する指定を与えるものであり、ENDのコマンドは検索終了を指示するものである。

第5図において、先ず処理100でキーボード

30からコマンドを入力すると、処理200ではそのコマンドを解析し、その内容がSTORE、FIND、DELETEのコマンドのときは、それぞれ処理300、400、500を実行する。コマンドがENDコマンドのときは、全処理を終了する。

第6図は、第5図の処理300、つまりSTOREコマンドの実行手順を示すフローチャートである。

先ず、処理310では、登録すべき文書の文書名、キーワードをキーボード30から読み込む。処理320では、画像入力装置70から上記文書を読み込み、画像ディスプレイ80に表示する。処理330では、文書の内容、例えばテキスト、画像、図表等によつて領域分割を行う。領域分割は画像であれば、読み込まれたデータ上に黒の部分が大きく、高周波成分が強いのに対して、図表であれば白い部分が大きく、黒ランレングスが短い等の特徴を用いて実行することができる。なお、人間が、表示された文書を見ながら指示することにより、分割を行うことも可能である。領域分割を行うときのデータ・タイプは、この例では、テキ

スト、画像、図表としたが、これらは目的等に応じて自由に設定できる。領域分割の結果は、画像ディスプレイ80に表示する。処理340では、分割後の各領域のデータを、領域ごとに文書ファイルに蓄積する。処理350では、当該文書の文書名、キーワード、文書中の各領域のデータ・タイプ、位置、大きさ、文書ファイル上のデータ番号を、検索データ・ベースに登録する。

第7図は、第5図の処理400、つまりFINDコマンドの実行フローチャートである。

先ず、処理410では、検索すべき文書の検索情報をキーボード30より入力する。検索情報としては、文書名、キーワード、領域に関する情報等を単独、または複合して指定できる。処理420では、検索データ・ベースより指定された検索情報に合致する文書名を求める。処理430では、指定された検索情報に合致する文書数に応じて、下記の各処理に進む。すなわち、合致した文書数が0の場合には処理490に進み、指定の検索情報に合致する文書が文書ファイル内に存在しない

ことを示すメッセージを、画像ディスプレイに表示する。また、合致した文書数が1の場合、処理480に進み、文書ファイルから当該文書のデータを読み出して、画像ディスプレイ80に表示する。また、合致した文書数が1より多い場合、処理440に進み、各文書の領域に関する情報、データ・タイプ、位置、大きさを検索データ・ベースから読み出し、データ・タイプに応じたパターンで領域を画像ディスプレイ80に表示する。処理450では、キーボード30より、次の処理を決定するための指定を読み込む。処理460では、キーボード30から読み込んだ指示に応じて、次の処理に進む。すなわち、(i)領域の内容表示要求の場合、キーボード30からどの領域の内容を表示するかを指定することにより、処理470に進み、指定された領域のデータを文書ファイルから読み出して、画像ディスプレイ80に表示する。また(ii)次ページ表示要求の場合、処理440に戻り、次ページの文書の領域情報を読み出して表示する。(iii)目的文書を指定した場合、

処理480に進み、文書ファイルから当該文書のデータを読み出し、画像ディスプレイ80に表示する。

第8図は、第5図の処理500、つまりDELETEコマンドの実行フローチャートである。

まず、処理510では、削除したい文書の文書名をキーボード30から読み込み、処理520で、検索データ・ベース上の該当データを削除する。処理530では、文書ファイル上の該当データを削除する。

本発明の要点を列記すると、次の4つになる。

(i) 指定の検索情報を有する文書データがファイル内に複数個存在する場合、当該の候補データの各領域を、そのデータ・タイプに応じたパターンで表示することにより、文書の構造を概略で表示することができ、またパターンで表示されている領域を指定して、その内容を見ることができるので、候補データの中から目的のデータを簡単に探索することができる。

(ii) 領域をパターンで表示する場合、検索デ

タ・ベース内の領域に関するデータ・タイプ、位置、大きさのみの情報で表示させることができ、かつ文書ファイルから画像データを読み出す必要がないため、きわめて高速に表示することができる。また、見たい領域のみを選択して表示することもできるため、効率のよい検索が可能となる。

(iii) 検索を行う場合、文書の構造に関する情報、例えば「図が二つ存在する。」、「右下に写真がある。」等を指定して、検索を行うことができ、従来では見られない一種の内容検索が可能である。

(iv) 検索の過程に応じて、領域のパターン表示のみによる文書表示から、特定の領域、例えば右下の画像領域等を指定した文書まで、概略表示から詳細表示までの柔軟な表示が可能であるため、効率的な検索を行うことができる。

#### (発明の効果)

以上説明したように、本発明によれば、文書の構造情報を利用して検索を行うので、操作性に優れ、かつ高速表示が可能で、効率のよい検索を行うことができる。

#### 図面の簡単な説明

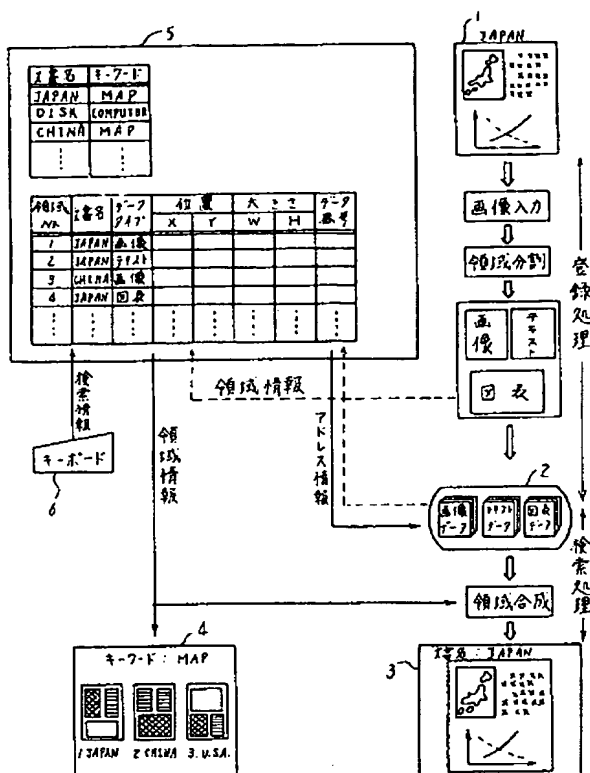
第1図は本発明の原理を示す概念図、第2図は第1図の文書と領域の位置関係を示す図、第3図は本発明の一実施例を示す文書ファイル検索システムのブロック図、第4図は本発明に用いるコマンドと機能の説明図、第5図は第4図のコマンドの処理フローチャート、第6図、第7図、第8図はそれぞれ第5図のSTORE、FIND、DELETEのコマンドの処理フローチャートである。

1…文書、2…文書ファイル、3…文書表示画面、4…構造表示画面、5…検索データ・ベース、6、30…キーボード、10…マイクロプロセッサ、20…主メモリ、40…ファイル装置、50…画像プロセッサ、60…画像メモリ、70…画像入力装置、80…画像ディスプレイ、90…データ・ベース・システム。

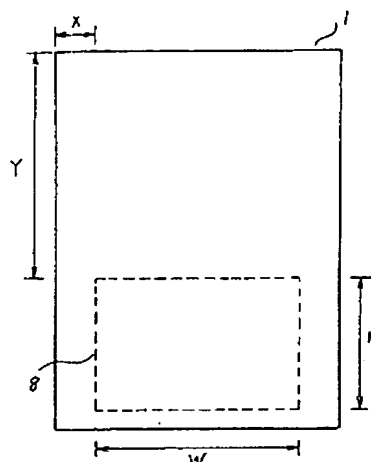
代理人 弁理士 高橋明



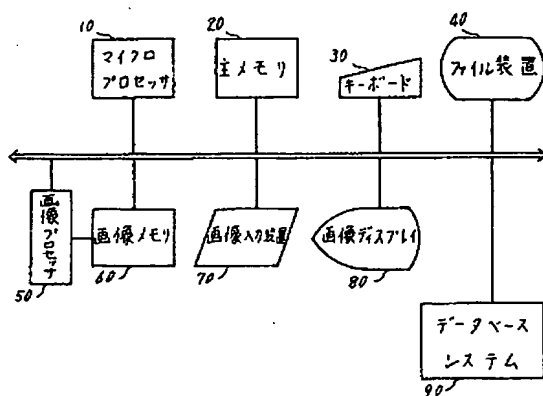
第 1 図



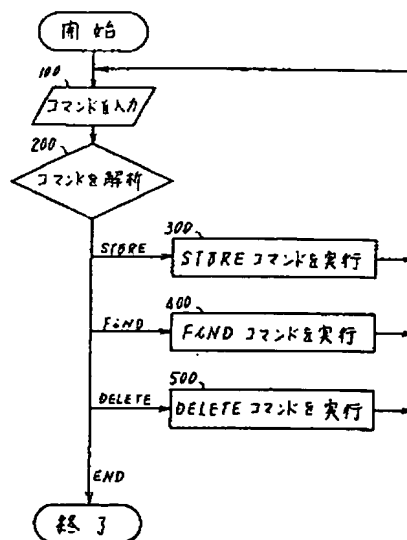
第 2 図



第 3 図



第 5 図

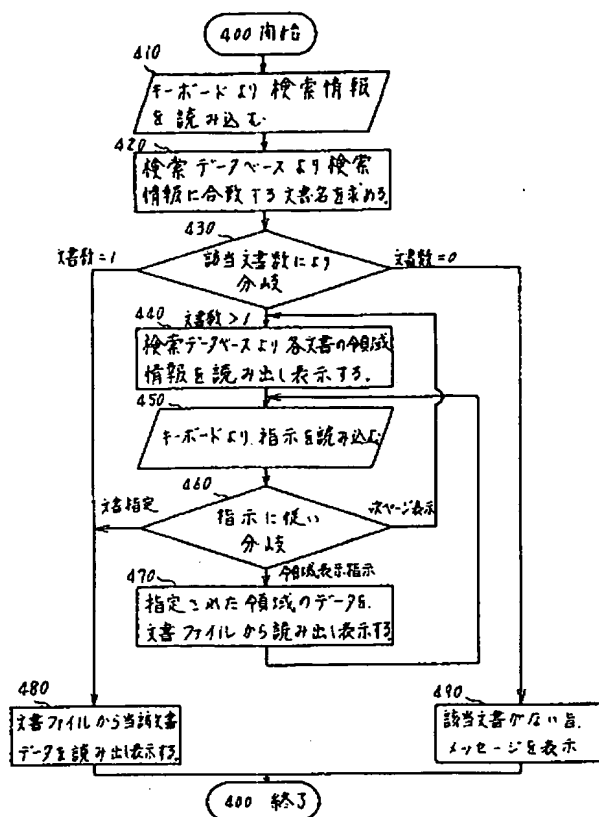
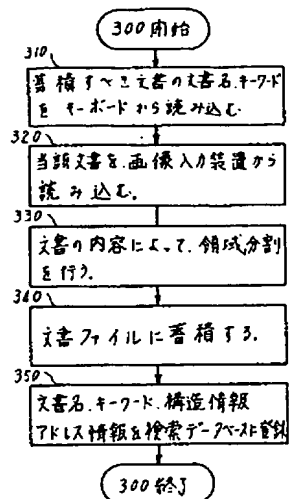


第 4 図

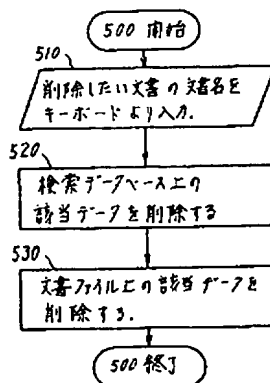
コマンド	機 能
STORE	文書と画像入力装置から読み込み領域分割して、検索データベースと書込みファイルに登録蓄積
FIND	指定の検索情報に合致する文書と検索データベースに照合検索し、文書ファイルの読み出し画像ディスプレイに表示。
DELETE	指定の文書データと書込みファイルの削除。
END	検索処理を終了。

第 7 図

第 6 図



第 8 図





**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**